Quality Control of Weather Data during Extreme Events

JINSHENG YOU AND KENNETH G. HUBBARD

High Plains Regional Climate Center, University of Nebraska at Lincoln, Lincoln, Nebraska

(Manuscript received 17 March 2005, in final form 4 August 2005)

ABSTRACT

Quality assurance (QA) procedures have been automated to reduce the time and labor necessary to discover outliers in weather data. Measurements from neighboring stations are used in this study in a spatial regression test to provide preliminary estimates of the measured data points. The new method does not assign the largest weight to the nearest estimate but, instead, assigns the weights according to the standard error of estimate. In this paper, the spatial test was employed to study patterns in flagged data in the following extreme events: the 1993 Midwest floods, the 2002 drought, Hurricane Andrew (1992), and a series of cold fronts during October 1990. The location of flagged records and the influence zones for such events relative to QA were compared. The behavior of the spatial test in these events provides important information on the probability of making a type I error in the assignment of the quality control flag. Simple pattern recognition tools that identify zones wherein frequent flagging occurs are illustrated. These tools serve as a means of resetting QA flags to minimize the number of type I errors as demonstrated for the extreme events included here.

1. Introduction

Quality assurance (QA) procedures have been applied (Guttman and Quayle 1990) to (semi)automatically check the validity of weather data from National Oceanic and Atmospheric Administration (NOAA) Climatological Observer stations archived at the National Climate Data Center (NCDC). While NCDC's processes of validation became semiautomated, there continued to be a major role for human data validators whose performance was assessed objectively (Guttman et al. 1988).

General testing approaches such as the threshold method and the step change test were designed for the single-station review of data to detect potential outliers (Wade 1987; Meek and Hatfield 1994; Eischeid et al. 1995). Quality assurance procedures based on physical processes have less history than the traditional statistical procedures. Examples include the testing of hourly solar radiation against the clear-sky envelope (Allen 1996; Geiger et al. 2002) and the use of soil heat diffusion theory to determine soil temperature validity (Hu et al. 2002).

Recently, the use of multiple stations in QA procedures has proven useful; for example, the spatial tests compare a station's data against the data from neighboring stations (Wade 1987; Gandin 1988; Eischeid et al. 1995; Eischeid et al. 2000; Hubbard 2001; Feng et al. 2004). The spatial tests involve the use of neighboring stations to make an estimate of the measurement at the station of interest. This estimate can be formed by weighting according to the inverse of the distance to the location (Guttman et al. 1988; Wade 1987), or through other statistical approaches [e.g., multiple regression, Eischeid et al. (1995) and Eischeid et al. (2000); bivariate linear regression test, Hubbard et al. (2005)].

The spatial regression test (SRT) used herein (Hubbard et al. 2005) does not assign the largest weight to the nearest neighbor but, instead, assigns the weights according to the standard error of estimate between the station of interest and each neighboring station. Hubbard et al. (2005) used seeded errors to test the performance of the threshold method, the step change method, the persistence test, and the spatial regression test. It was found that the spatial regression test outperformed the other three methods, which missed many of the errors identified by the spatial regression

Corresponding author address: Jinsheng You, High Plains Regional Climate Center, University of Nebraska at Lincoln, Lincoln, NE 68583-0728.
E-mail: jyou2@unl.edu

test while identifying only a handful beyond those identified by the spatial regression test.

In a separate study, a sensitivity analysis was performed on various user-selected parameters for implementation of the spatial regression test (Hubbard and You 2005). The sensitivity analysis included the response to the radius of inclusion, the regression time window, the regression time offset, and the number of stations used to make the estimates. The performance of the SRT method was found to be stable when 10 or more stations were applied in the estimates.

The earlier study demonstrated the excellent performance of the spatial regression test in identifying the seeded errors. Experiments are still needed to answer such questions as how will the spatial regression test perform in extreme but real events, especially for data collected during an extreme wet year or an extreme dry year, in which precipitation at some stations may experience record or near-record daily values, monthly values, or both? In this study we investigate the performance of the spatial regression test for such extreme events as the 2002 U.S. drought (the extreme dry/hot condition), the 1993 Midwest floods (the extremely wet condition), Hurricane Andrew (1992), and a series of cold-air outbreaks in October 1990. These events allow us to characterize the behavior of the spatial regression test in extreme events and to explore causes and patterns of the QA failures (QA failure is defined here as a higher frequency of flagging at a station than is consistent with the frequency of flags at other stations in the region; this is a QA failure in the sense that manual intervention is required to actually determine if the data are good or bad).

2. Methods and data

In this study, we used the data from both the Cooperative Observer Weather Network and the Automated Weather Data Network (AWDN), and other weather data networks such as the hourly surface airways network and the Historical Climatology Network. We conducted only one QA iteration regardless of the number of potentially incorrect data entries identified. Currently the NCDC and other regional climate centers archive the $T_{\max}$ and $T_{\min}$ in degrees Fahrenheit. To be consistent with the widespread use of this data in Fahrenheit and consistent with the database, we use degrees Fahrenheit in this paper.

a. Quality assurance rules

Quality assurance software consists of procedures or rules against which data are tested. Each procedure will either accept the data as being true or reject the data and label it as a potential outlier. This is hypothesis (Ho) testing of the data and the statistical decision to accept the data or to note it as a potential outlier can have the outcomes listed in Table 1.

Take the simple case of testing a variable against limits. If we take as our hypothesis that the data for a measured variable are valid only if they lie within $\pm 3\sigma$ of the mean (μ), then assuming a normal distribution we expect to accept Ho 99.73% of the time with no error. The values that lie beyond $\mu \pm 3\sigma$ will be rejected and we will make a type I error when we encounter valid values beyond these limits. In this case, we are rejecting Ho when the value is actually valid and we expect to make a type I error 0.27% of the time. This is true for data that have no errant values. If an "errant" value is encountered, then the hypothesis will properly be rejected, only if the errant value falls outside the range $\mu \pm 3\sigma$. It would otherwise be accepted, when it is actually false (the value is not valid) and result in a type II error. In this simple example, reducing the limits against which the data values are tested will produce more type I errors and fewer type II errors while increasing the limits leads to fewer type I errors and more type II errors. For quality assurance software, tuning is necessary to achieve a balance wherein one reduces the type II errors (mark less errant data as having passed the test) while not increasing type I errors to the point where valid extremes are brought into question. Because type I errors cannot be avoided, it is essential for data managers to always keep the original measured values regardless of the quality testing results and further examine those values that have been flagged.

b. Spatial regression test

The spatial regression test (Hubbard et al. 2005) is a quality control approach that checks whether the variable falls inside the confidence interval formed from the measurements at surrounding stations during a time period of length n (60 days in this study). All stations (M) within a certain distance of the station of interest are selected and a linear regression is performed for each station paired with the station of interest and centered on the datum of interest. For each surrounding

TABLE 1. Hypothesis rules.

Statistical decision	True situation: Ho true	True situation: Ho false
Accept Ho	No error	Type II error
Reject Ho	Type I error	No error

station, a regression-based estimate ($x_i = a_i + b_i y_i$) is formed. The weighted estimate (x') is obtained by utilizing the standard error of estimate (s) also known as rmse in the weighting process:

$$x' = \pm\left[\sum_{i=1}^{N} \pm x_i^2 s_i^{-2}\right]^{0.5}\left[\sum_{i=1}^{N} s_i^{-2}\right]^{-0.5}. \qquad (1)$$

Here, N is the number of stations to be used in the estimate (generally restricted to an R^2 greater than 0.5 within a given radius of inclusion; in this study we use an inclusion of 1°). Note that N is selected by the user and $N \leq M$. Care must be taken to preserve the correct sign on x' and x_i. The negative sign is used inside $\Sigma \pm x_i^2 s_i^{-2}$ when x_i is negative. When $\Sigma \pm x_i^2 s_i^{-2}$ is negative, the negative sign is used for the estimated value and outside $\Sigma \pm x_i^2 s_i^{-2}$ to maintain the validity of the equation. The weighted standard error of estimate (s') is calculated as

$$\frac{1}{s'^2} = N^{-1}\sum_{i=1}^{N} s_i^{-2}. \qquad (2)$$

Now the confidence intervals can be based on s' and we test whether or not the station value (x) falls within the confidence intervals defined by f:

$$x' - fs' \leq x \leq x' + fs'. \qquad (3)$$

If the value of x in (3) causes the relation in Eq. (3) to be true, then the corresponding data pass the spatial regression test. This relationship indicates that increasing f decreases the number of potential type I errors. Unlike distance weighting techniques, this approach does not assume that the nearest station will get the largest weighting.

c. *Extreme events*

The tests conducted in this study were carried out over a wide spatial range during extreme or major events like flooding, droughts, tropical storms, and cold fronts. The data from stations were retrieved through the Applied Climate Information System (ACIS), a distributed data management system (Hubbard et al. 2004). The data sources include the measurements from the NOAA Cooperative Observer Weather Network, the Automated Weather Data Network [AWDN: the Automated Surface Observing System (ASOS), the Automated Weather Observation System (AWOS)], and other networks such as the hourly surface airways network and the Historical Climatology Network. This study includes the extreme events selected for quality assurance of variables like the maximum air temperature, the minimum air temperature, and the precipitation. The selected events and the related variables are shown in Table 2.

The Missouri River basin and the upper Mississippi River basin experienced record heavy rain and flooding in the summer of 1993. This event was included in the analyses and precipitation was the variable examined in the quality control (QC) analysis. In 2002, 48% of the United States was in a drought at the end of August. This event was included as the drought event and maximum temperature is the variable that we discuss in the QC analysis. Hurricane Andrew (1992) devastated Dade County, Florida, with an estimated $25 billion in damage after its first landfall at Biscayne Bay on 24 August 1992. It made a second landfall in south-central Louisiana on August 26. We examine how many of the large precipitation amounts associated with the passage of the hurricane were flagged as outliers. More detailed information of these events can be found in the NCDC database of extreme events (NCDC 2005).

A series of cold front events moved across the continental United States in October of 1990 and the passages of the cold front events were well documented on the daily weather maps. The air temperature and precipitation are affected by the passage of these cold fronts, so T_{max}, T_{min}, and precipitation were all examined in the QC analysis. The goal is to understand how the spatial regression test will perform in the transition associated with these weather patterns.

3. Results and discussion

In each of the following tests the number of flags and their spatial and temporal disposition were tracked. A discussion of the patterns created by flags is provided.

TABLE 2. Selected data and events.

		Variables		
Locations	Events	T_{max}	T_{min}	Precipitation
Midwest (Missouri watershed and upper Mississippi)	Extreme wet events (1993)			Yes
Western United States	Extreme dry event (2002)	Yes		
Coastal southeastern United States	Hurricane Andrew (1992)			Yes
Continental United States	Oct cold fronts (1990)	Yes	Yes	Yes

a. Relationship between interval of measurement and QA failures

Analyses were conducted to prepare artificial maximum and minimum temperature values (not the measurements, but the values identified as the maximum and minimum from the hourly time series) for different times of observation from available hourly time series of measurements. The observation time for coop weather stations varies from site to site. Here we define the A.M. station, P.M. station, and nighttime station according to the time of observation as listed in Table 3. The cooperative network generally uses the P.M. measurements with a significant number of A.M. measurements, and the Automated Weather Data Network uses a midnight-to-midnight observation period.

The daily precipitation accumulates the precipitation for the past 24 h ending at the time of observation. The precipitation during the time interval may not match the precipitation from nearby neighboring stations due to event slicing; that is, precipitation may occur both before and after a station's time of observation. In Table 3 we specify that the precipitation observations, regardless of time of observations, provide good measurements because the mass is conserved for different times of observation.

The measurements of the maximum and the minimum temperature are the result of making discrete intervals on a continuous variable. The maximum or minimum temperature takes the maximum value or the minimum value of the temperature during the specific time interval. Thus, the maximum temperature or the minimum temperature is not necessarily the maximum or minimum value of a diurnal cycle. Figure 1 shows an example of these differences obtained from three time intervals. The hourly measurements of air temperature were retrieved from 0100 local time (LT) 11 March to 1700 LT 13 March 2002 at Mitchell, Nebraska. The times of observation are marked: "A" shows the minimum air temperature obtained for 11 March for A.M. stations, and "B" is the maximum temperature obtained for 13 March at the P.M. stations. The minimum temperature may carry over to the following interval

TABLE 3. Time interval and possible performance of three intervals of measurements.

	A.M. station	P.M. station	Nighttime station (AWDN)
Time intervals (e.g.)	~0700	~1700	~0000
Max temperature		Problematic	
Min temperature	Problematic		
Precipitation	Good	Good	Good

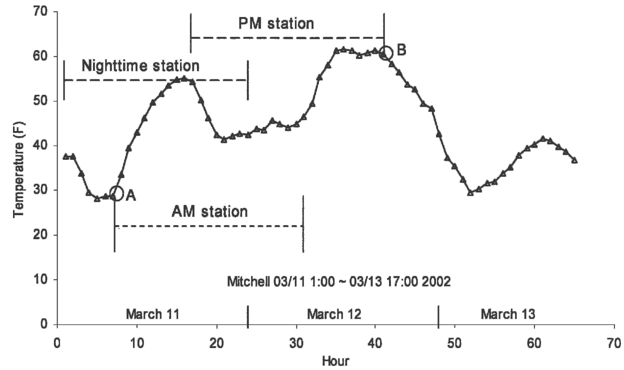

FIG. 1. Exemplary time intervals of measurements at Mitchell, NE. The "A" denotes the 11 Mar minimum temperature measured at the A.M. station. The "B" denotes the 13 Mar maximum temperature at the P.M. station.

for A.M. stations and the maximum temperature may carry over to the following interval for P.M. stations. We have therefore marked these as problematic in Table 3 to note that the thermodynamic state of the atmosphere will be represented differently for A.M. and P.M. stations.

Figure 2 gives the time series of (a) the daily precipitation, (b) the daily maximum air temperature, and (c) the daily minimum air temperature. The circles delineate those measurements obtained at the P.M. station that have a high risk of QA failure when compared to neighboring A.M. stations. The bias of the daily precipitation obtained using different times of observation can be significant and reaches 2.5 in. day^{-1} in this example, which will cause a flag unless adjustments or limited comparisons are employed. Similar cases exist for the maximum and minimum temperature, where temperature differences reach 20°F between stations with different time intervals. Here, we note that the QA failures are not due to sensor problems but to comparing data from stations where the sensors are employed differently. To avoid this problem A.M. stations can be compared to A.M. stations, P.M. stations to P.M. stations, etc. In this study we still compare all stations together because we wish to see whether the time shifting and dynamically derived station weights of the SRT can handle the situations with mixed observation times. Note that this problem will be solved if modernization of the coop network provides hourly or subhourly data at most station sites.

b. 1993 floods

Quality control procedures were applied to the data for the 1993 Midwest floods over the Missouri River basin and part of the upper Mississippi River basin,

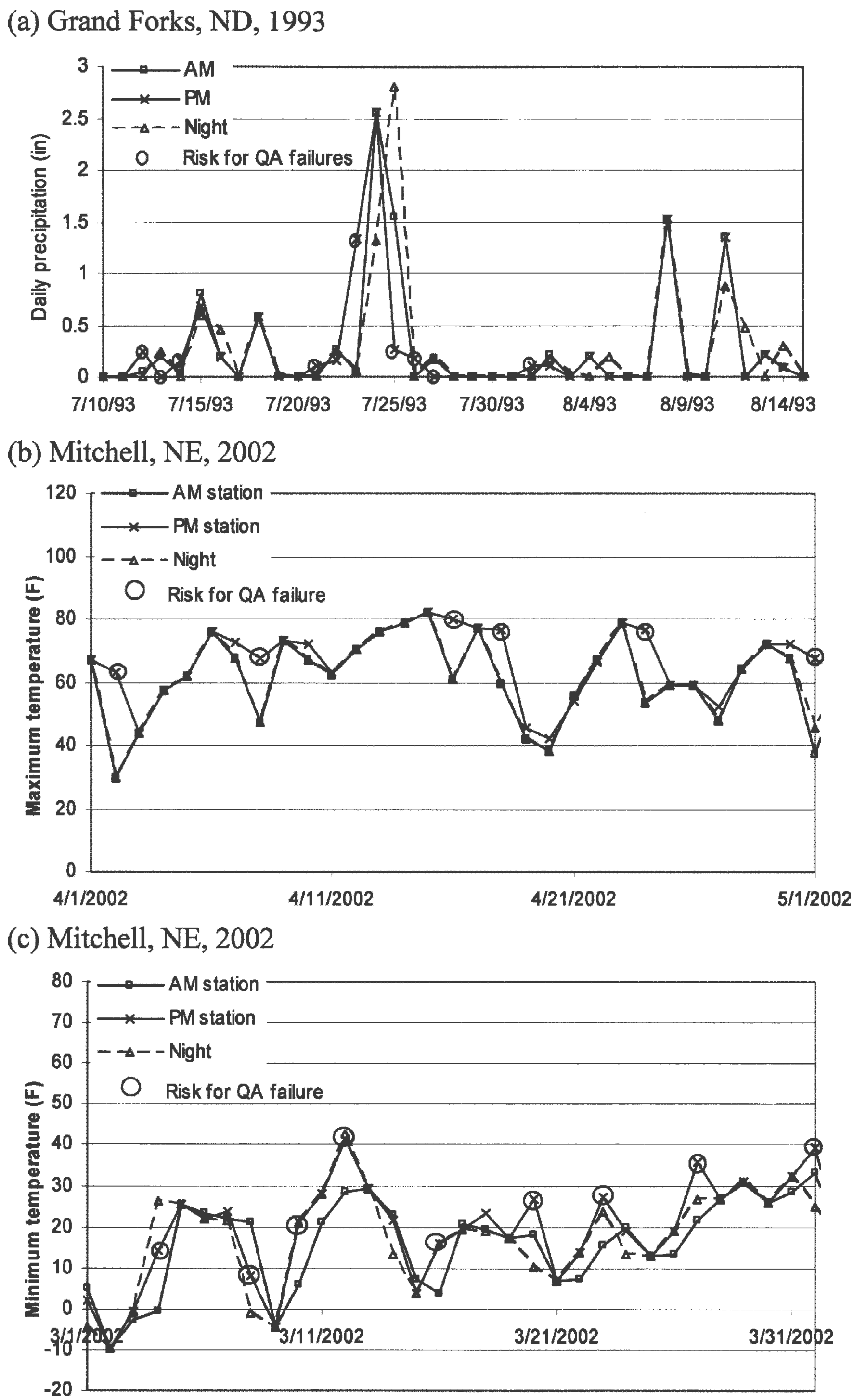

FIG. 2. Differences between measurement intervals of A.M. stations and P.M. stations lead to the risk of QA failures of (a) precipitation, (b) maximum air temperature, and (c) minimum air temperature.

where heavy rainfall and floods occurred. Data were tested from 1 April 1 to 30 August. Figure 3 shows the interpolated spatial pattern of the fraction of flagged precipitation records for stations during the 1993 Midwest floods. The spatial regression test performs well and flags 5% ~ 7% of the data for most of the area at f = 3. We note that the heavy precipitation events were fairly uniform as seen by return periods plotted in

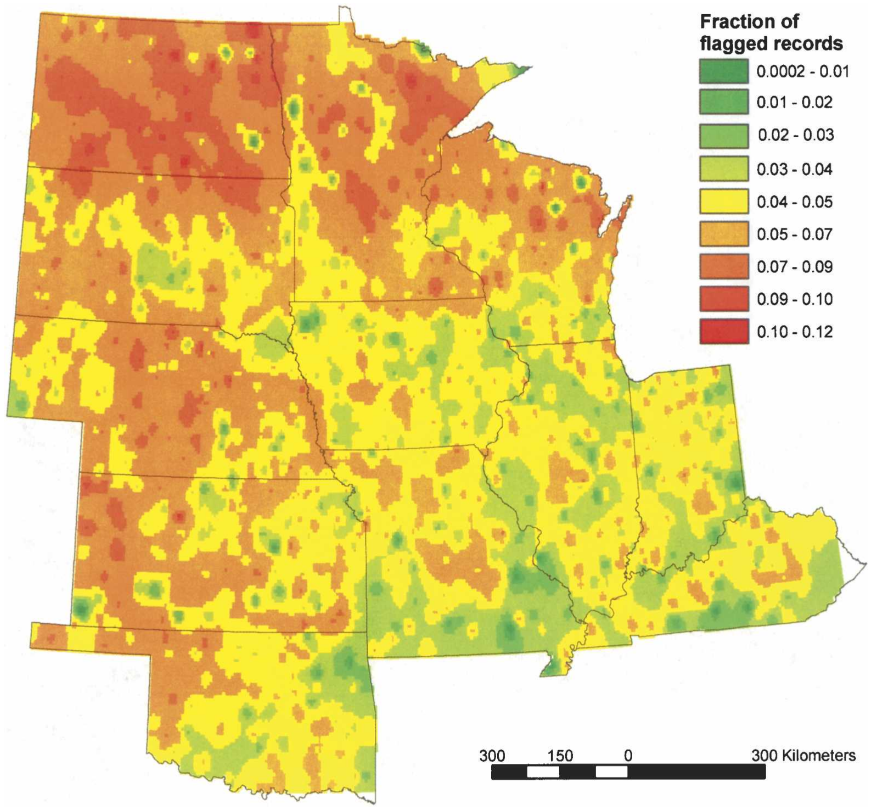

FIG. 3. Interpolated spatial pattern of fractions of flagged precipitation records at $f = 3$.

Fig. 4 (Lott 1994). The spatial patterns of the fraction of the flagged records do not coincide with the spatial pattern of the return period. For example, the southeastern part of Nebraska does not show a high fraction of flagged records although most stations have return periods of more than 1000 yr. Also, northern Wisconsin has a higher fraction of flagged records while the precipitation for this case has a lower return period in that area.

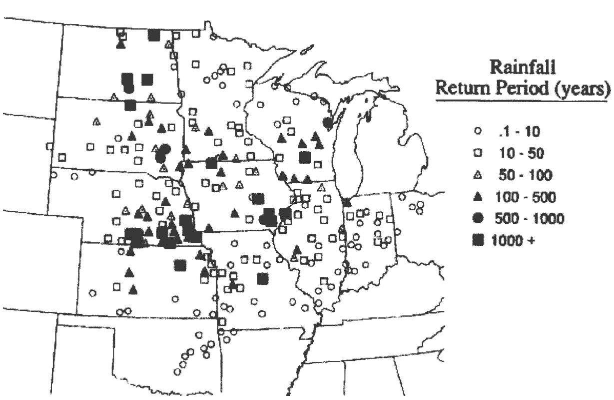

FIG. 4. Rainfall return period (yr) for rainfall amounts that fell during June–July 1993, for the upper Midwest area (after Lott 1994).

A more detailed analysis was applied to the state of North Dakota. North Dakota has a significantly higher fraction of flagged records than other states. The analysis of the Grand Forks AWDN station demonstrates that the differences in daily precipitation obtained from stations with different times of observation contributed to the high fraction of QA failures. Figure 2a demonstrates that nine records had a high risk of failure during the time period from 1 July to 15 August 1993 when the measurements of the current station and the reference station are obtained from P.M. and A.M. stations, respectively. The situation worsens if the measurements at weather stations were obtained from different time intervals and the distribution of stations with different times of observation is unfavorable. This would be the case for an isolated A.M. or P.M. station.

Among the 13 flags at Grand Forks, 9 may be due to the different times of observation or perhaps the size and spacing of clouds. Four other flags occurred during localized precipitation events, in which only a single station received significant precipitation. For example, Grand Forks had 1.52 in. of precipitation on 8 August,

TABLE 4. Number of stations with the indicated number of flags (NOFs) and the fraction.

NOF	0	1	2	3	4	5	6	7	8	9	10
No.	1275	1107	680	314	133	62	24	7	3	1	1
FOFR	0.3535	0.3069	0.1885	0.0871	0.0369	0.0172	0.0067	0.0019	0.0008	0.0003	0.0003

FOFR: fraction of flagged records.

while the other stations had less precipitation with the highest being 0.84 in. However, the 2-day (8 and 9 August) precipitation for the other stations ranges from 0.39 to 2.74 in. compared with 1.53 in. at Grand Forks. Higher precipitation entries occurring in isolation are more likely to be identified as potential outliers. We expect that some of these problems can be avoided by examining the precipitation over larger intervals, for example, summing consecutive days into event totals.

c. 2002 drought events

Tests were conducted over the western United States from 1 March though 30 August 2002. The number of potential outliers is not significant for the maximum air temperature. Among 3607 stations, only one had a number of flags (NOF) equal to 10 while another station had an NOF equal 9. A total of 1275 stations do not have flagged records and 1107 stations have only one flagged record. Table 4 lists the frequency of flags and how many stations reached this frequency. Figure 5 shows the locations of the stations with flagged records.

Table 5 is an example of one of these stations (Agate, Nebraska). Most of the flagged values were cases where the measurement was higher than those at surrounding stations. Through analysis of the hourly data from the neighboring AWDN station at Mitchell, Nebraska (Fig. 2b), we find that only one flag, which occurred on 16 August 2002, was caused by the bias in the regression models. All other flags (nine in total) were caused by differences of the values obtained for different times of observation. Here, we list an example of the spatial

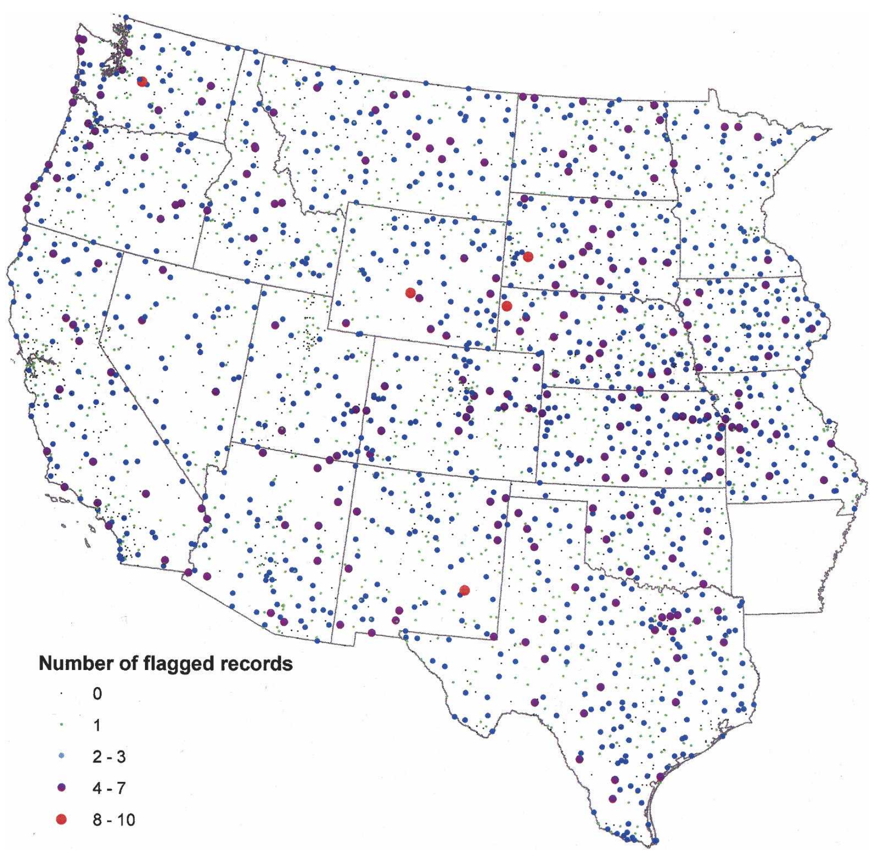

FIG. 5. Spatial pattern of number of flagged records during the 2002 drought event in the western United States.

TABLE 5. Flagged records at meteorological station 250030 Agate 3E, NE.

Date	T_{max}	Flag	Estimated	Date	Tmax	Flag	Estimated
6 Mar 2002	51	Failed	35.11	21 Jul 2002	94	Failed	83.07
13 Mar 2002	55	Failed	39.88	1 Aug 2002	97	Failed	80.38
2 Apr 2002	46	Failed	29.51	12 Aug 2002	92	Failed	72.18
8 Apr 2002	68	Failed	47.39	16 Aug 2002	101	Failed	89.38
24 Apr 2002	75	Failed	52.45	17 Aug 2002	100	Failed	86.04

regression estimate for Agate on 6 March. The spatial regression test at the Agate station uses the data from six neighboring stations and the values used in the regression on 6 March 2002 are 23°, 33°, 27°, 36°, and 44.7°F, which gave an estimate of 35.11°, 16°F lower than the measurement. From the daily time series we would say this is a good measurement given that the measurements from 4 March to 8 March are 45.0°, 55.0°, 51.0°, 55.0°, and 22.0°F. Thus, this is a type I error.

Besides the five highlighted stations, Fig. 5 also shows the pattern of the flagged records as a fraction of the total records. No significant relationship is found between the topography and the fraction of flagged records. Some clusters of stations with high flag frequency are located along the mountains; however, other mountainous stations do not show this pattern. Moreover, some locations with similar topographical setting have different patterns. For the state of Colorado, a high fraction of flags occurs along the foothills of the Rocky Mountains where the mountains meet the high plains. A high fraction was also found along Interstate Highways 25 and 70 in east Colorado (Fig. 6). Along Interstate 70, one station has six flagged records and two stations have four flagged records, and most stations have two or more flagged records. However, most stations do not have flagged records in the regions to the north or to the south of Interstate 70. A similar pattern can be identified along Interstate 25, which is located along the seam that joins the Rocky Mountains

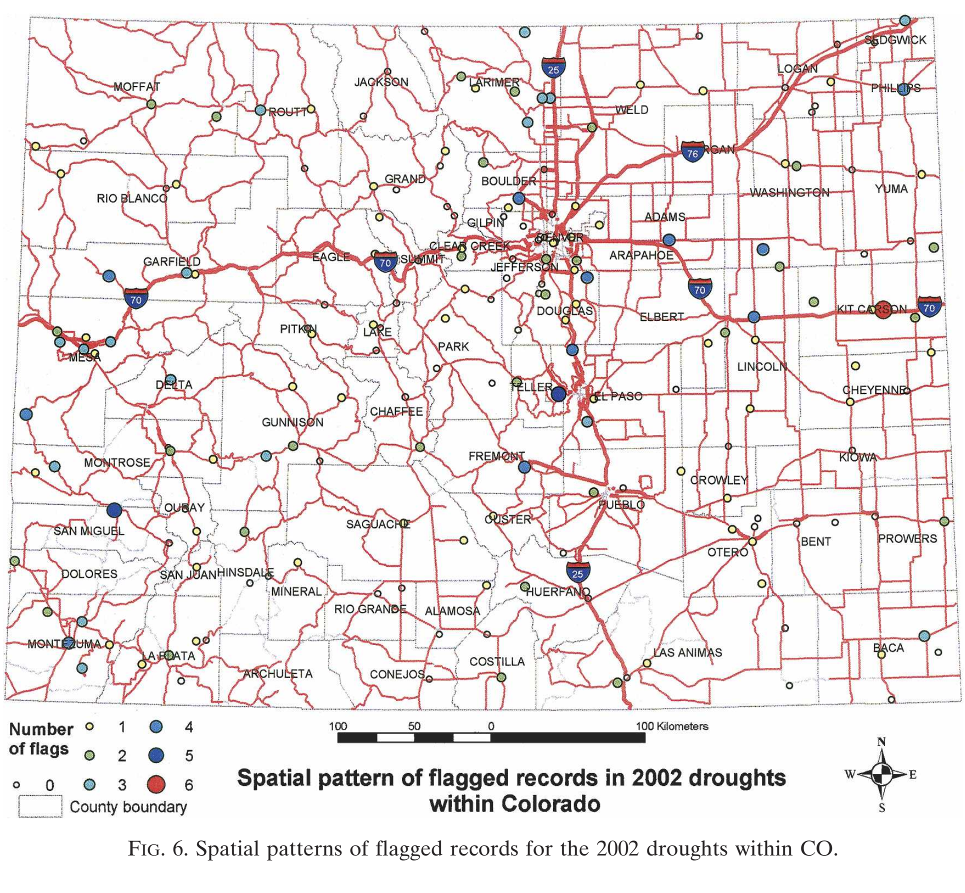

FIG. 6. Spatial patterns of flagged records for the 2002 droughts within CO.

and the high plains. We were not able to determine if the higher number of flags along the interstates was a statistical anomaly that perhaps occurred only during the 2002 drought or if there is a physical reason (e.g., management, instrumentation, etc.) for the situation.

Instrumental failures and abnormal events also lead to QA failures. Figure 7 shows the time series of Stratton Station in Colorado, operated as part of the automated weather network. This station has nighttime (midnight) readings, while all of the neighboring sites are A.M. or P.M. stations. Stratton thus has the most flagged records in the state (six) and the highlighted records were flagged. We checked the hourly data time series to investigate the QA failure in the daily maximum temperature time series for the time period from 20 April to 20 May 2002. No value was found to support a T_{max} of 88° for 6 May in the hourly time series; thus, 88°F appears to be an outlier. On 7 May, a high of 85°F is recorded for the P.M. station observation interval, in which the value of the afternoon of 6 May is recorded as the high on 7 May. The 102°F observation of 8 May at 0600 LT appears to be an observation error caused by a spike in the instrument reading. The observation of 93°F at 0800 LT 17 May is supported by the hourly observation time series (see Fig. 7b) and is apparently associated with a downburst from a decaying thunderstorm.

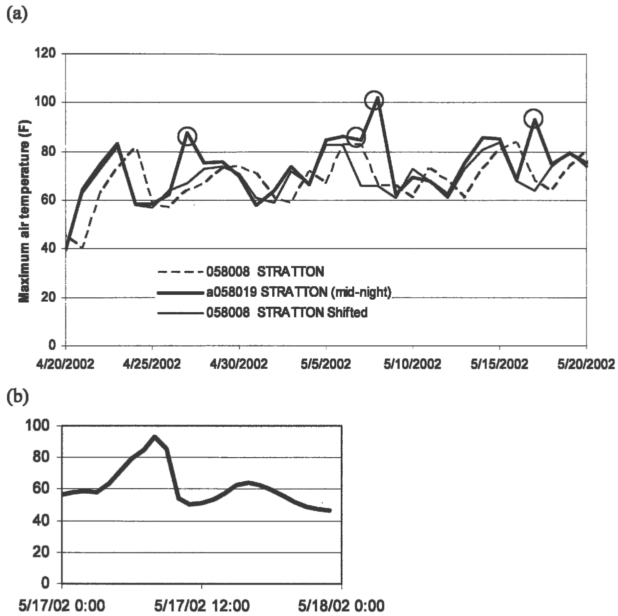

FIG. 7. Time series of Stratton and a neighboring station during the 2002 droughts: (a) the daily time series of T_{max} for Stratton and the Stratton AWDN station (a058019) and (b) the hourly time series at the Stratton AWDN station.

d. 1992 Andrew Hurricane

Figure 8 shows the evolution of the spatial pattern of flagged records from 25 to 28 August 1992 during Hurricane Andrew and the corresponding daily weather maps. The flags in the spatial pattern figures are cumulative for the days indicated. The test shows that the spatial regression test explicitly marks the track of the tropical storm. Starting from the second landfall of Hurricane Andrew at central-southern Louisiana, the weather stations along the route have flagged records. The wind field formed by Hurricane Andrew helps to define the influence zone of the hurricane on flags. Many stations without flags have daily precipitation of more than 2 in. as the hurricane passes, which confirms that the spatial regression test is performing reasonably well.

e. Cold front 1990

The cold front events during October 1990 were tested and only the results of the maximum air temperature on 6 and 7 October are shown (Fig. 9). The maximum air temperature dropped by as much as 40°F during the passage of the cold front. Spatial patterns of flags on 6 October coincide with the area traversed by the cold front and many stations were flagged in such states as North Dakota, South Dakota, Iowa, and Nebraska. On 7 October, the cold front moved to southeast regions beyond Nebraska and Iowa. Of course, nearby stations on opposite sides of the cold front may experience different temperatures thus leading to flags. This may be further complicated when different times of observation are involved. The cold front continues moving and the flags also move with the front correspondingly.

A similar phenomenon can be found in the test of the precipitation and the minimum temperature. A spatial regression test of any of these three variables can roughly mark the movements of the cold front events. The identified movements of the cold fronts and associated flagging of "good records" may lead to more manual work to examine the records. However, if we understand the cause of the flags, we may be able to employ a reverse flagging scheme to minimize the problem of flagging "good data." Simple pattern recognition tools have been developed to identify the spatial patterns of these flags and reset these flags automatically.

The spatial patterns of flagged records are significant for both the spatial regression tests of the cold front events and the tropical storm events. However, most of these flagged records are type I errors; thus, we developed pattern recognition tools to assist in reducing these flags. Differences still exist between the distribu-

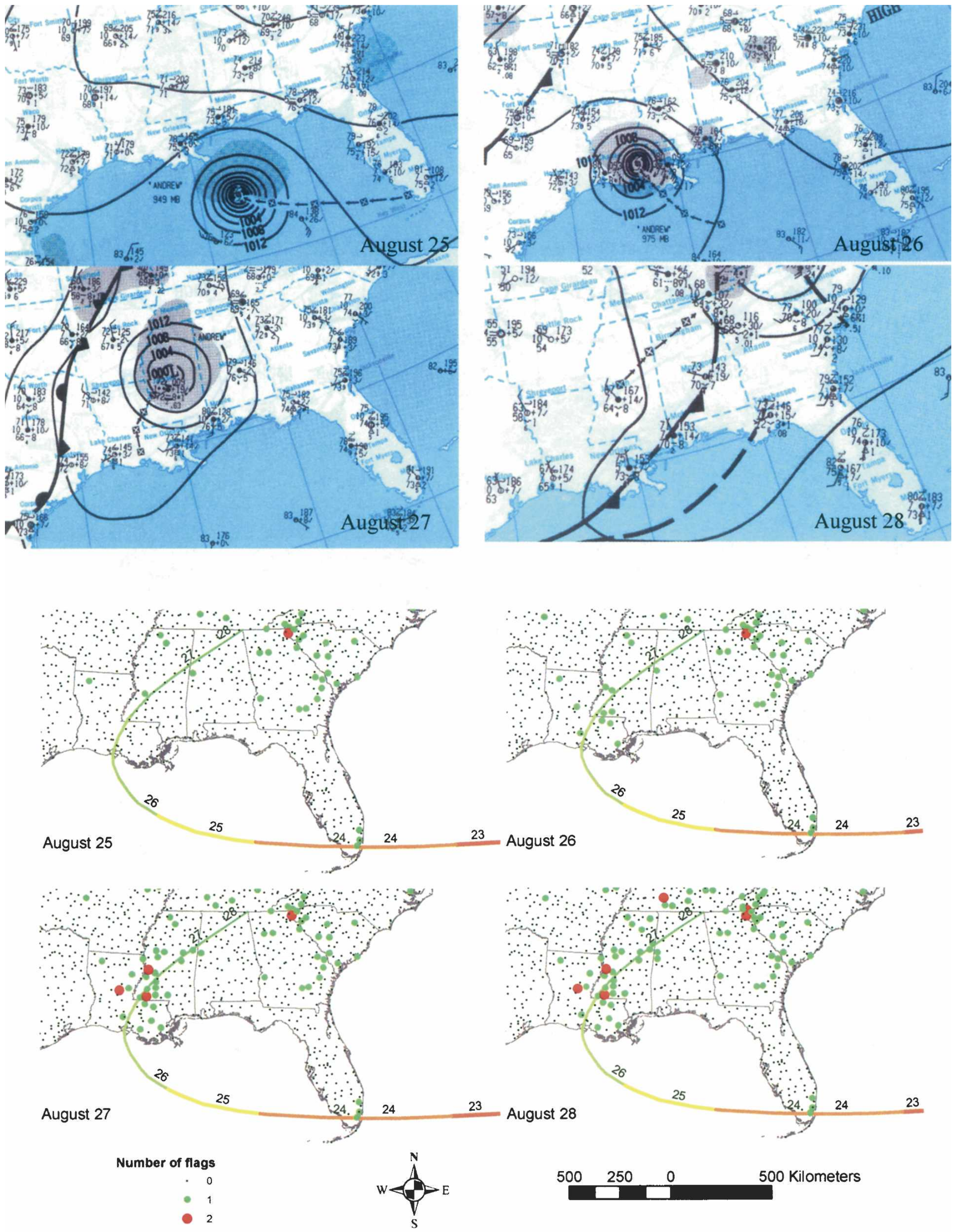

FIG. 8. Daily weather maps and spatial pattern of flagged records for the 1992 Andrew Hurricane events.

tion patterns of the flagged records for the cold front event and the tropical storm events due to the characteristics of cold front events and tropical storm events. These differences include the following.

1) The influence zone of the cold front relative to QC is determined by how rapidly the front advances from day to day. Thus, stations may be switched in a 24-h period from observing on the warm side of the front to observing on the cold side. Some cold fronts move rapidly thus creating a large influence zone in their wake where a significant number of stations may be flagged. The boundary between no rain or light rain and heavy rain determines the zone of

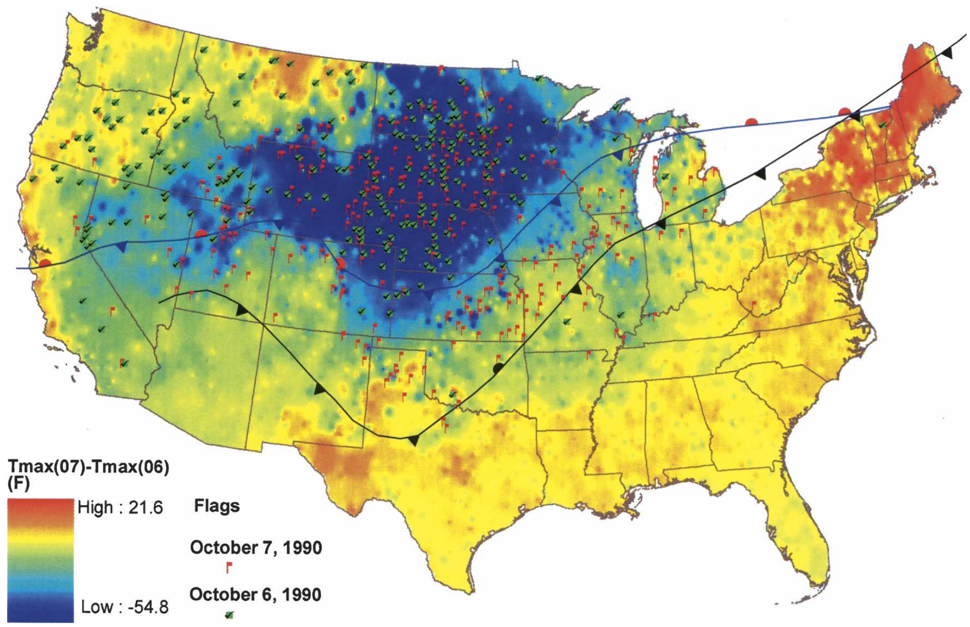

FIG. 9. Spatial patterns of flagged records for cold front events and related fronts. The temperature map is the interpolated maximum temperature difference between 6 and 7 Oct 1990. The colored front is on 7 Oct, and the black one is on 6 Oct. The flags are the QA failures on that day.

influence on QC in tropical storms. The influence zone of the tropical storms are smaller and only the stations along the storm route and the neighboring stations have flags.

2) Cold fronts exert influences on both the air temperature and precipitation. The temperature differences between the regions immediately ahead of the cold fronts and regions behind can reach 20° ~ 40°F. The precipitation events caused by the cold fronts may be significant, depending on the moisture in the atmosphere during the passage. The tropical storms generally produce a significant amount of precipitation. A few inches of rainfall in 24 h is very common along the track because the tropical storms generally carry a large amount of moisture.

f. Reset the flags for cold front events and hurricanes

Some measurements during the cold fronts and the hurricane were valid but flagged as outliers due to the effect on QC tests of the considerable temperature changes caused by the cold front passages and the heavy precipitation occurring in hurricanes. We developed a simple spatial scheme to recognize regions where flags have been set due to type I errors. The stations along the cold front may experience the mixing situation where some stations have been affected by the cold fronts and others have not. A complex pattern recognition method can be applied to identify the influence zone of the cold fronts through the temperature changes [e.g., using some methods described in Jain et al. (2000)]. In this paper, we use the simple rule to reset the flag given that significant temperature changes occur when the cold front passes. The mean and the standard deviation of the temperature change can be calculated as

$$\overline{\Delta T} = \frac{1}{n}\sum_{i=1}^{n} \Delta T_i \quad \text{and} \tag{4}$$

$$\sigma^2_{\Delta T} = \frac{1}{n}\sum_{i=0}^{n} (\Delta T_i^* \Delta T_i) - \overline{\Delta T}^* \overline{\Delta T}, \tag{5}$$

where $\overline{\Delta T}$ is the mean temperature change of the reference stations, ΔT_i is the temperature change at the ith station for the current day, n is the number of neighboring stations, and $\sigma_{\Delta T}$ is the standard deviation of the temperature change for the current day. A second round test is applied to records that were flagged in the first round:

$$\overline{\Delta T} - f'\sigma_{\Delta T} \le \Delta T \le \overline{\Delta T} + f'\sigma_{\Delta T}, \tag{6}$$

where ΔT is the difference between maximum or minimum air temperature for the current day and the last day. The cutoff value f' takes a value of 3.0. The test results for the $T_{\max}$ with this refinement are shown in Fig. 10 for 7 October 1990. The results obtained using

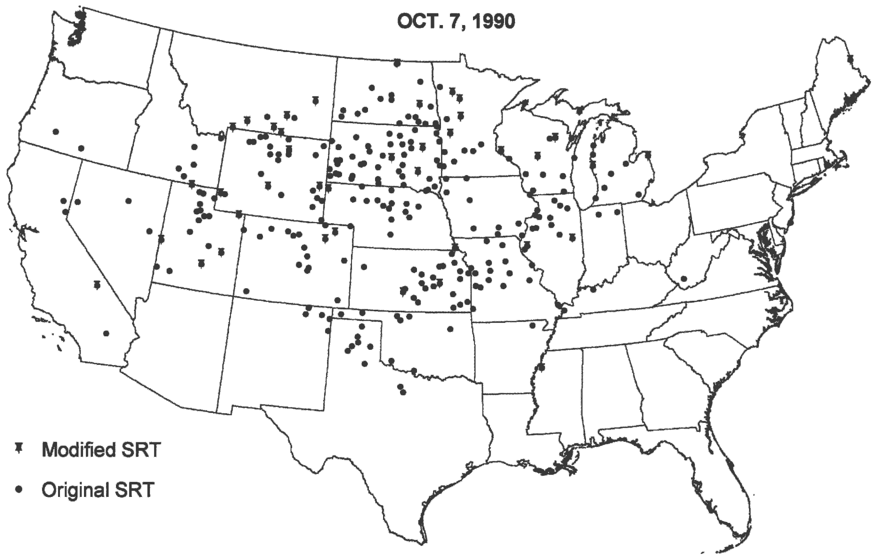

FIG. 10. All points shown were flagged by the original SRT method, while the thumbtacks were those that are flagged by the modified SRT method for $T_{\max}$. The filled circles are those that are reset by the modified SRT method.

the refinements described in this section were labeled "modified SRT" and the results using the original SRT were labeled "original SRT" in Figs. 10 and 11. Of the 291 flags originally noted only 41 flags remain after the reset phase. The daily temperature drops more than 20°F at most stations where the flags were reset and the largest drop is 55°F.

For the heavy precipitation events, we compare the amount of precipitation at neighboring stations to see whether heavy precipitation occurred. We use a similar approach as for temperature to check the number of neighboring stations that have significant precipitation:

$$\zeta = \text{count}(p_i \geq p_{\text{threshold}}), \tag{7}$$

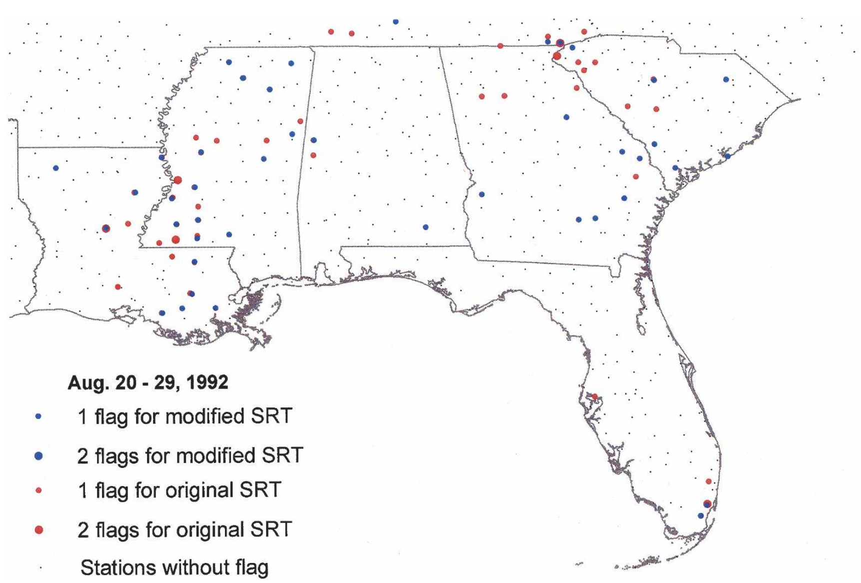

FIG. 11. Reset flags for Hurricane Andrew. The flags are the cumulative flags starting from 20 to 29 Aug 1992. The flags by the modified SRT method overlay the flags by the original SRT method.

where the p_i is the daily precipitation amount at a neighboring station and $p_{\text{threshold}}$ is a threshold beyond which we recognize that a significant precipitation event has occurred at the neighboring station, for example, 1 in. When $\zeta \geq 2$ and $p < p_{\text{high}}$, we reset the flag. Here, p is the current precipitation amount at the station, and p_{high} is the upper threshold at the station of interest beyond which the threshold test will flag the measurement. For this study p_{high} was taken as the value for which precipitation is exceeded only 97.5% of the time. Figure 11 shows maps of flags after the reset process. Of the 78 flags originally noted, only 41 flags remain after the reset phase. Most of the remaining flags are due to the precipitation being higher than the upper threshold, which are type I errors of the test.

4. Discussion and conclusions

The spatial regression test was applied to a number of representative extreme wet and dry events. This technique was successful in that the quality assurance procedures for the extreme wet or extreme dry condition did not produce a significantly higher fraction of flagged records than were produced during nonextreme conditions of wetness and dryness. The spatial regression test can be applied in the quality control procedures of the extreme events; however, this approach still flagged some extreme readings that may be valid new records. The isolated rainfall events are far more problematic than the uniform wet conditions and the uneven spatial distribution of the weather variables leads to more flags. Without further QC tools, the flags produced will need to be examined manually.

The spatial regression test does not show a significant pattern related to the extreme wet condition or the drought events. Some patterns of frequent flagging were observed in the passage of cold front and tropical storm situations. Some flagged records of maximum temperature were caused by the different observation times. The P.M. stations may observe higher maximum temperature than the A.M. stations after a significant high. These values do not conform to a diurnal cycle even though the observations may be accurate and recorded correctly. Instrument failures also lead to abnormal measurements and these records may be flagged in the QA procedure.

In these tests, many flags are actually type I errors. Most of these errors can be avoided by increasing the validity zone of the test (increasing f). However, this measure increases the risk of producing more type II errors while reducing the type I errors. The spatial regression test is robust in identifying the extreme events, which helps the automatic quality assurance procedures, especially for the automatic weather data network.

We have developed a test to recognize the zones where frequent flagging occurred. By identifying the spatial patterns of the flags in the extreme events, flags associated with type I errors can be safely reset. In this paper, the mean and the standard deviation of temperature changes of the reference stations were applied to discriminate whether a cold front event affects the station of interest. The test indicates that the procedures can successfully reset the flags. The modified SRT approach works well for all the cold events when adopting the simple statistical approach to reset the flags. The flags for precipitation associated with the hurricanes can also be reset by referring to the precipitation events at neighboring stations. Further research to optimize the performance of these methods is under way.

REFERENCES

Allen, R. G., 1996: Assessing integrity of weather data for reference evapotranspiration estimation. *J. Irrig. Drainage Eng.*, **122,** 97–106.

Eischeid, J. K., C. B. Baker, T. Karl, and H. F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.,* **34,** 2787–2795.

——, P. A. Pasteris, H. F. Diaz, M. S. Plantico, and N. J. Lott, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteor.,* **39,** 1580–1591.

Feng, S., Q. Hu, and W. Wang, 2004: Quality control of daily meteorological data in China 1951–2000: A new dataset. *Int. J. Climatol.,* **24,** 853–870.

Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.,* **116,** 1137–1156.

Geiger, M., L. Diabate, L. Menard, and L. Wald, 2002: A Web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy,* **73,** 475–480.

Guttman, N. V., and R. G. Quayle, 1990: A review of cooperative temperature data validation. *J. Atmos. Oceanic Technol.,* **7,** 334–339.

——, C. Karl, T. Reek, and V. Shuler, 1988: Measuring the performance of data validators. *Bull. Amer. Meteor. Soc.,* **69,** 1448–1452.

Hu, Q., S. Feng, and G. Schaefer, 2002: Quality control for USDA NRCS SM-ST network soil temperatures: A method and dataset. *J. Appl. Meteor.,* **41,** 607–619.

Hubbard, K. G., 2001: Automated Weather Stations for Applications in Agriculture and Water Resources Management. AGM-3 WMO/TD No. 1074, 248 pp.

——, and J. You, 2005: Sensitivity analysis of quality assurance using a spatial regression approach—A case study of the maximum/minimum air temperature. *J. Atmos. Oceanic Technol.,* **22,** 1520–1530.

——, A. T. DeGaetano, and K. D. Robbins, 2004: Announcing a modern Applied Climatic Information System (ACIS). *Bull. Amer. Meteor. Soc.,* **85,** 811–812.

——, S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi,

2005: Performance of quality assurance procedures for an Applied Climate Information System. *J. Atmos. Oceanic Technol.,* **22,** 105–112.

Jain, A. K., R. R. W. Duin, and J. Mao, 2000: Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.,* **22,** 4–37.

Lott, J. N., 1994: The U.S. summer of 1993: A sharp contrast in weather extremes. *Weather,* **49** (11), 370–383.

Meek, D. W., and J. L. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteor.,* **69,** 85–109.

NCDC, cited 2005: Extreme weather and climate events. [Available online at http://www.ncdc.noaa.gov/oa/climate/severeweather/extremes.html.]

Wade, C. G., 1987: A quality control program for surface mesometeorological data. *J. Atmos. Oceanic Technol.,* **4,** 435–453.